Supply Chain Risk Assessment Prediction Using Machine Learning

Saish Milind Sail

B.E. Mechanical Engineering, Birla Institute of Technology and Science, Pilani – Goa Campus

25th April, 2025

Abstract
Supply chain disruptions pose significant risks to businesses, affecting operational efficiency and financial stability. This study leverages machine learning techniques to predict key disruption-related factors, including penalty costs, compensation paid, disruption costs, and time to recovery. Various regression models were implemented and evaluated to determine their effectiveness in forecasting these variables. This study adopts a structured methodology comprising data preprocessing, feature engineering, model evaluation, hyperparameter tuning, and SHAP-based explainability. Each phase is designed to enhance the accuracy and interpretability of risk predictions. A Streamlit dashboard was developed to enable real-time inference, user interaction, and visual explanations of model outputs. By integrating predictive modelling with interpretable AI, the system empowers businesses to make proactive, data-driven decisions, reduce potential losses, and strengthen supply chain resilience.

1. Introduction

Supply chains are complex networks susceptible to disruptions caused by logistics delays, supplier failures, and external events. Accurate prediction of disruption-related factors is essential for risk mitigation and financial planning. This study explores machine learning-based approaches to predicting key supply chain variables. The primary objectives include:

1. Developing a structured dataset for supply chain disruption analysis.
2. Implementing machine learning models to predict disruption-related financial impacts.
3. Evaluating model performance using RMSE and R^2 scores.

2. Literature Review:

The paper by Jahin et al. (2024) examines the role of Artificial Intelligence (AI) and Machine Learning (ML) in enhancing Supply Chain Risk Assessment (SCRA). Traditional risk management methods are increasingly inadequate in handling the complexities and uncertainties of modern supply chains, especially post-COVID. This review addresses a research gap by exploring emerging AI/ML techniques like Random Forest and XGBoost, which significantly improve predictive accuracy and risk mitigation strategies.

The study reviews 1,439 papers, focusing on 51 articles published between 2015 and 2024, and provides key insights into AI/ML's impact on SCRA. It highlights how these techniques have led to more resilient contingency plans and adaptive risk management strategies. The paper also includes a bibliometric analysis, revealing trends, influential authors, and highly cited works, offering valuable insights into the field's growth.

In conclusion, the review provides a comprehensive roadmap for future research and practical applications, emphasizing AI/ML's transformative potential in fortifying supply chain risk management.

3. Data Collection and Preprocessing

3.1 Data Acquisition

The dataset was generated through a combination of web scraping from supply chain-related databases and synthetic data augmentation using SMOTE. Key sources include supplier reliability reports, logistics performance indices, and financial impact assessments from disruptions.

3.2 Dataset Overview

The dataset includes various attributes relevant to supply chain disruptions, such as:

- **Supplier_Reliability_Score**: A numerical representation of supplier performance.
- **On_Time_Delivery_Rate**: The percentage of orders a supplier delivers on time. A higher value means the supplier is reliable, while a lower value indicates frequent delays.
- **Financial_Stability_Score**: A score from 0 to 100 indicating a supplier's financial health. A higher score means the supplier is financially strong, while a lower score suggests a risk of bankruptcy or disruptions.
- **Severity**: The level of impact caused by a disruption. It is categorized as Low, Medium, or High, where High severity disruptions have greater financial and operational consequences. The dataset uses predefined thresholds for classification:
    - Low Severity: Minimal financial loss (<$50,000) and recovery within a week.
    - Medium Severity: Moderate financial impact ($50,000–$500,000) with recovery in weeks to a few months.
    - High Severity: Major disruptions costing >$500,000 and requiring several months to recover.
- **Logistics_Delay**: Categorical feature indicating transportation delays.
- **Time_To_Recovery_Days**: The estimated number of days required to recover from a disruption. Higher values indicate longer recovery times, often seen in severe disruptions like natural disasters or bankruptcies.
- **Disruption Cost (USD):** Estimated financial loss due to a disruption (higher for high-severity events).
- **Insurance Coverage (%):** Percentage of disruption cost covered by insurance.
- **Penalty Cost (USD):** Extra cost due to late deliveries or non-compliance with contracts.
- **Compensation Paid (USD):** If the supplier had to compensate clients due to a major issue.

3.3 Data Preprocessing

To ensure data consistency, preprocessing steps included:

1. **Handling Missing Values**: Imputed using statistical methods.
2. **Encoding Categorical Variables**: One-hot encoding and label encoding for categorical data.
3. **Feature Scaling**: MinMax scaling applied to numerical features.
4. **Train-Test Split**: Data split into training (80%) and testing (20%) sets.

4. Machine Learning Models Implemented

The study employed multiple regression models to predict financial and recovery-related variables:

- **RMSE (Root Mean Squared Error)**
  - RMSE represents the average error in the same scale as the target variable. Since we applied MinMax scaling (0-1), the RMSE values tell us how far the predictions deviate from the actual values within the scaled range.
- **R^2 (Coefficient of Determination)**
  - R^2 measures how well the model explains variance in the target variable. $R^2 < 0$ means the model is worse than just predicting the mean for every instance. Ideally, R^2 should be closer to 1 for a good model.

4.1 Linear Regression

Linear Regression serves as a baseline model to understand the direct relationships between independent variables and target financial indicators. It assumes a linear relationship and minimizes error by fitting a straight-line equation to the dataset.

Pros:

- Simple and interpretable.
- Computationally efficient.

Cons:

- Assumes linearity, which may not hold for complex supply chain disruptions.
- Sensitive to outliers.

Results:

- Penalty_Cost_USD - RMSE: 0.1634, R^2 Score: -0.1001
- Compensation_Paid_USD - RMSE: 0.1709, R^2 Score: -0.5562
- Disruption_Cost_USD - RMSE: 0.2354, R^2 Score: -0.0343
- Time_To_Recovery_Days - RMSE: 0.3534, R^2 Score: -0.5211

Analysis:

An RMSE of 0.1634 (for Penalty Cost) means the model's typical prediction error is about 16.34% of the scaled range, which is moderate but not great.

Time_To_Recovery_Days has the worst RMSE at 0.3534, meaning the model's predictions are off by ~35% on average.

Compensation_Paid_USD (R^2 = -0.5562) → The worst-performing model; it's not capturing useful patterns.

Penalty_Cost_USD (R^2 = -0.1001) → The least bad, but still not reliable.

4.2 Decision Tree Regression

Decision Tree Regression builds a tree-like model of decisions, splitting data at different thresholds to minimize prediction errors. It captures non-linear relationships more effectively than Linear Regression.

Pros:

- Captures non-linear patterns.
- Handles categorical data efficiently.

Cons:

- Prone to overfitting.
- Sensitive to small variations in data.

Results:

- Penalty_Cost_USD - RMSE: 0.1382, R^2 Score: 0.2134
- Compensation_Paid_USD - RMSE: 0.1056, R^2 Score: 0.4063
- Disruption_Cost_USD - RMSE: 0.1764, R^2 Score: 0.4194
- Time_To_Recovery_Days - RMSE: 0.2859, R^2 Score: 0.0041

Analysis:

Penalty_Cost_USD - The model has improved over Linear Regression (which had R^2 = -0.1001), meaning it's capturing more patterns in the data.

However, an R^2 of 0.2134 suggests that the model explains only ~21% of the variance, which is still quite low.

Compensation_Paid_USD - This is the best-performing model so far, with 40.63% variance explained.

The RMSE is also the lowest among all target variables, indicating better predictions.

Disruption_Cost_USD - This model performs slightly better than for Compensation_Paid_USD, explaining 41.94% of the variance.

Decision Trees seem to be capturing patterns in Disruption Cost better than Linear Regression.

Time_To_Recovery_Days- The R^2 score is nearly 0, meaning the model is barely better than predicting the mean of the data.

The high RMSE suggests that Decision Trees are not effectively capturing patterns in this variable.

4.3 XGBoost Regression

XGBoost (Extreme Gradient Boosting) is an ensemble learning technique that improves prediction accuracy by combining multiple decision trees. It optimizes model performance using boosting techniques.

Pros:

- Highly accurate and efficient.
- Handles missing values effectively.

Cons:

- Requires hyperparameter tuning.
- Computationally expensive for large datasets.

Results:

- Penalty_Cost_USD - RMSE: 0.1025, R^2 Score: 0.5669
- Compensation_Paid_USD - RMSE: 0.0874, R^2 Score: 0.5933
- Disruption_Cost_USD - RMSE: 0.1380, R^2 Score: 0.6448
- Time_To_Recovery_Days - RMSE: 0.2191, R^2 Score: 0.4153

Analysis:

Compared to Decision Tree and Linear Regression, XGBoost performed significantly better in terms of R^2 score across all target variables.

RMSE values are lower than previous models, indicating that XGBoost produces more accurate predictions.

4.4 Random Forest Regression

Random Forest is an ensemble method that builds multiple decision trees and combines their outputs for improved prediction accuracy. It reduces variance and enhances generalization.

Pros:

- Robust against overfitting.
- Works well with both numerical and categorical data.

Cons:

- Slower compared to simple models.

- Less interpretable than individual decision trees.

Results:

- Penalty_Cost_USD - RMSE: 0.0990, R^2 Score: 0.5960
- Compensation_Paid_USD - RMSE: 0.0819, R^2 Score: 0.6424
- Disruption_Cost_USD - RMSE: 0.1343, R^2 Score: 0.6632
- Time_To_Recovery_Days - RMSE: 0.2243, R^2 Score: 0.3870

Analysis:

Both Random Forest and XGBoost perform significantly better than Linear and Decision Tree regression models. However, XGBoost still has a slight edge in terms of performance, especially for predicting Time_To_Recovery_Days.

5. Hyperparameter Tuning

To enhance model performance and ensure robust predictions, hyperparameter tuning was conducted for each of the four target variables: Penalty_Cost_USD, Compensation_Paid_USD, Disruption_Cost_USD, and Time_To_Recovery_Days. Both Random Forest and XGBoost regression models were tuned using GridSearchCV across relevant hyperparameter spaces. Performance metrics including Root Mean Squared Error (RMSE) and R^2 Score were used to assess model quality.

- **Penalty_Cost_USD**

For this monetary target, the Random Forest model achieved the best performance. The optimal hyperparameters were:

- max_depth = 10
- min_samples_leaf = 1
- min_samples_split = 5
- n_estimators = 200

This configuration yielded an RMSE of 0.0984 and an R^2 score of 0.6014. In comparison, XGBoost with the best settings (learning_rate = 0.05, max_depth = 3, n_estimators = 100, subsample = 1.0) performed slightly worse, with an RMSE of 0.1012 and R^2 of 0.5781.

- **Compensation_Paid_USD**

Similarly, for Compensation_Paid_USD, Random Forest outperformed XGBoost. The best parameters included:

- max_depth = 20
- min_samples_leaf = 2
- min_samples_split = 5
- n_estimators = 200

With these, Random Forest achieved an RMSE of 0.0788 and an R^2 score of 0.6688, compared to XGBoost's RMSE of 0.0838 and R^2 of 0.6255. Given the higher accuracy, Random Forest was selected for this target.

- **Disruption_Cost_USD**

In contrast, XGBoost emerged as the superior model for predicting Disruption_Cost_USD. The optimal XGBoost configuration (learning_rate = 0.05, max_depth = 3, n_estimators = 100, subsample = 1.0) achieved an RMSE of 0.1307 and an R^2 score of 0.6810. Random Forest trailed slightly with RMSE of 0.1318 and R^2 of 0.6759. Although the improvement was marginal, the consistent performance of XGBoost made it the preferred choice for this target.

- **Time_To_Recovery_Days**

Time_To_Recovery_Days is a critical operational metric where even small improvements can significantly affect planning. XGBoost again outperformed Random Forest, achieving:

RMSE: 0.2115

R^2: 0.4552

This was slightly better than Random Forest's performance (RMSE: 0.2183, R^2: 0.4195), making XGBoost the chosen model for this prediction task as well.

Model Selection Summary

Random Forest was selected for predicting Penalty_Cost_USD and Compensation_Paid_USD due to its stronger performance on financial variables.

XGBoost was selected for Disruption_Cost_USD and Time_To_Recovery_Days, where its precision provides an advantage for sensitive operational decisions.

6. SHAP implementation

To foster transparency and enhance trust in the model predictions, we incorporated SHAP (SHapley Additive exPlanations) for interpreting the output of the trained models. SHAP provides both global and local explanations, making it possible to understand how each feature contributes to a particular prediction or affects the overall model behavior.

Global Explainability

For each of the four target variables, global SHAP values were computed using the best-performing models identified during hyperparameter tuning. The summary plots generated highlight the most influential features across the entire dataset.

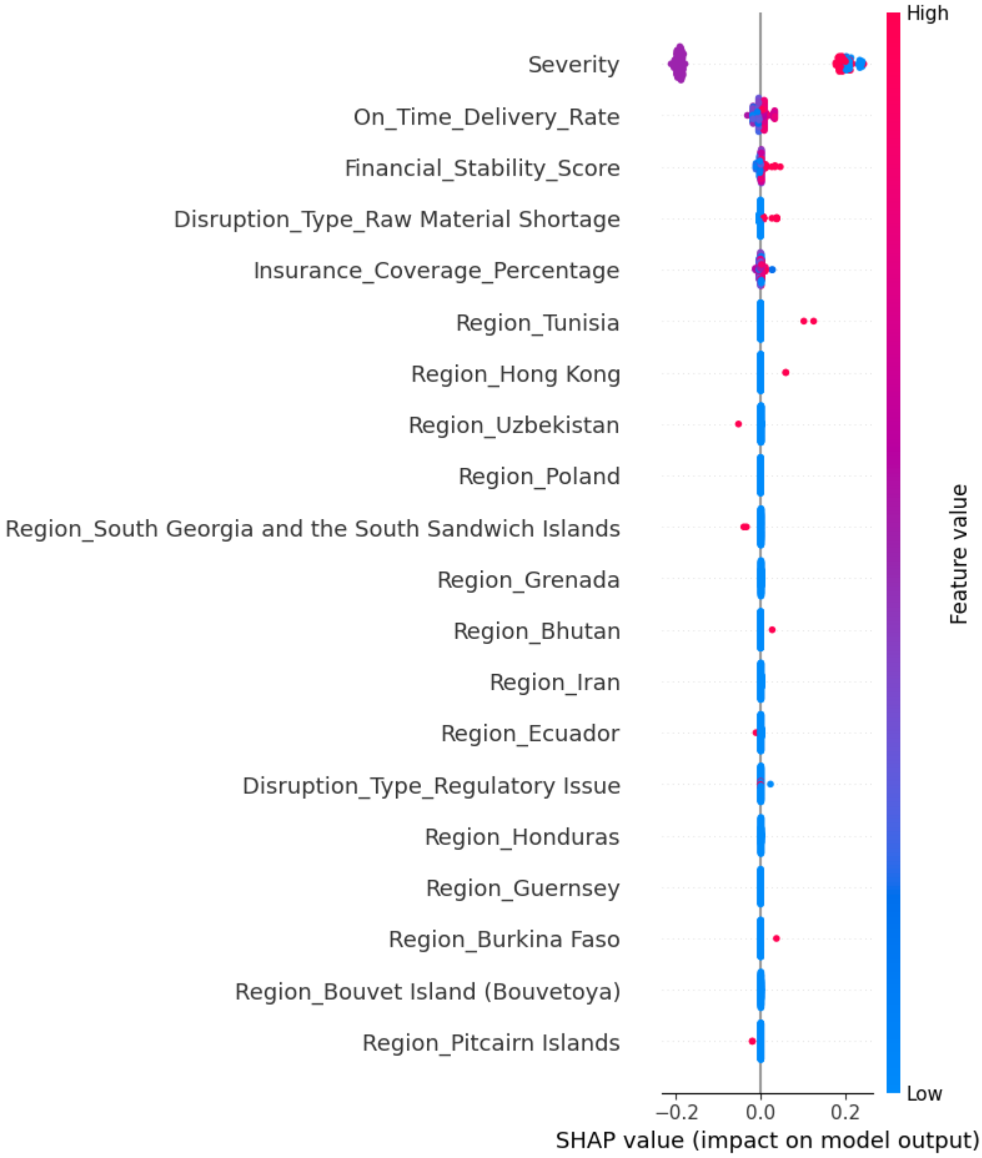
Local Explainability

To provide granular decision support, individual SHAP force plots were generated for selected supplier cases. These local explanations help stakeholders understand why a specific prediction was made, offering clarity in high-stakes scenarios such as evaluating penalty liabilities or planning compensation budgets.

6.1 Disruption_Cost_USD

- Severity is the most influential feature. Higher severity (shown in red) increases disruption cost significantly.
- Insurance_Coverage_Percentage and Financial_Stability_Score also impact the prediction, where lower values (in blue) tend to push the disruption cost higher.
- On_Time_Delivery_Rate has a moderate impact — lower delivery rates slightly increase the predicted cost.
- Certain Region features (like Greece, Cayman Islands, Colombia, etc.) have minor but visible effects, indicating some geographical influence.

6.2 Time_To_Recovery_Days

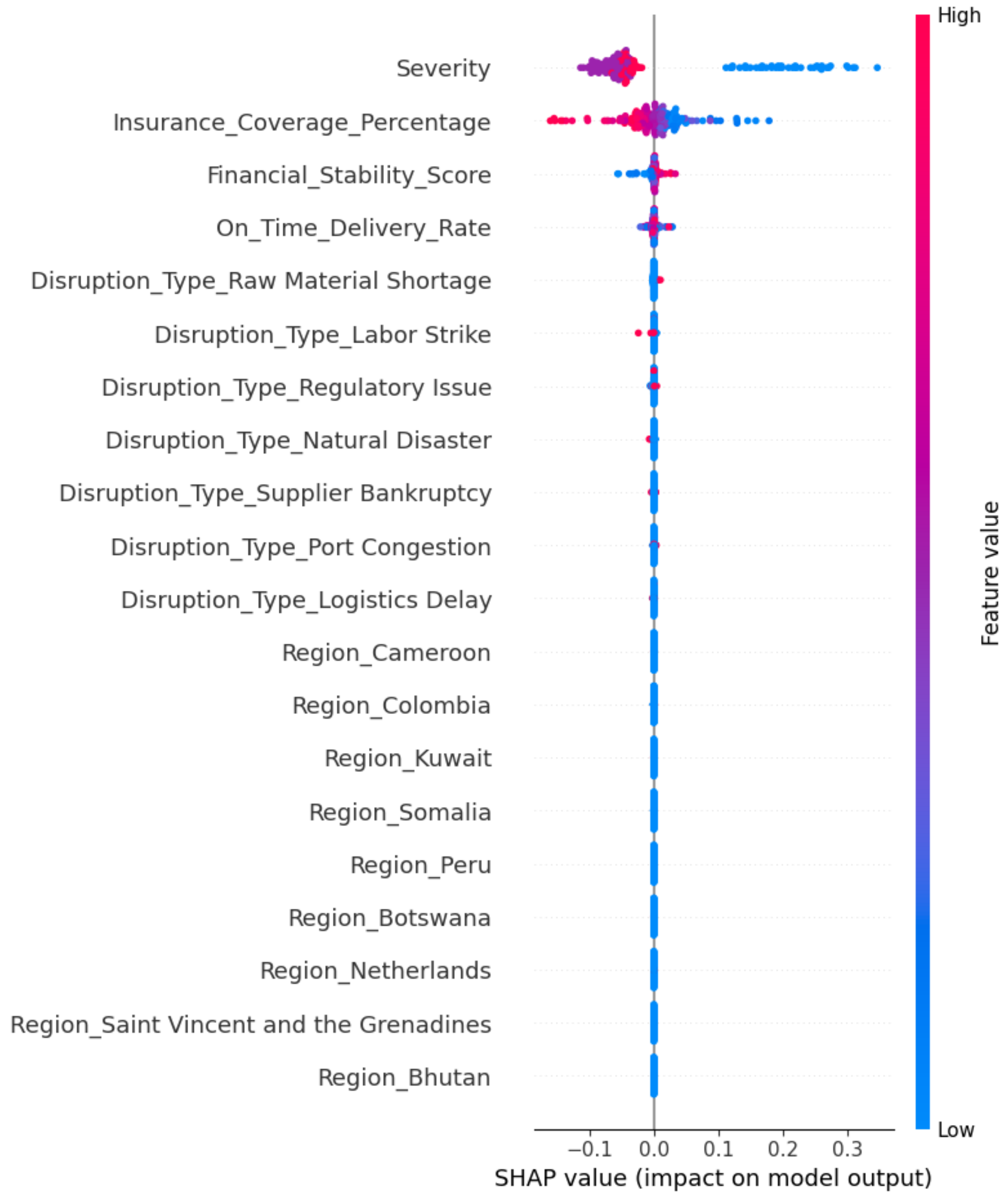


- Severity again stands out as the most important factor — higher severity levels (in red) tend to increase the recovery time.
- On_Time_Delivery_Rate and Financial_Stability_Score follow closely. Lower delivery rates and weaker financial scores push predictions toward longer recovery periods.
- Disruption_Type_Raw Material Shortage has noticeable impact — presence of this disruption type tends to increase recovery time.
- Insurance_Coverage_Percentage has a modest influence, where lower coverage slightly increases recovery days.
- Several regional features (e.g., Tunisia, Hong Kong, Uzbekistan) show limited but present effects.

6.3 Penalty_Cost_USD

- Severity is again the dominant factor — high severity leads to higher penalty costs.
- Insurance_Coverage_Percentage and On_Time_Delivery_Rate have strong influence too. Low insurance coverage and poor delivery rates push up penalties.
- Financial_Stability_Score contributes as well — financially unstable suppliers tend to be associated with higher penalty risk.
- Certain disruption types like Port Congestion and Supplier Bankruptcy also increase penalty likelihood.
- The remaining features, including several regions and disruption types (e.g., Raw Material Shortage, Natural Disaster), show minor effects but are still worth noting in specific scenarios.

6.4 Compensation_Paid_USD



- Severity continues to be the top driver — higher severity levels are clearly linked to higher compensation costs.
- Insurance_Coverage_Percentage is highly influential too — low coverage (in blue) leads to larger compensation liabilities.
- Financial_Stability_Score and On_Time_Delivery_Rate show that poor financial health and unreliable delivery contribute to higher compensation.
- Disruption types like Raw Material Shortage, Labor Strike, Natural Disaster, and Regulatory Issues all push compensation costs upward.
- The geographic regions at the bottom of the chart contribute relatively little, indicating that disruption type and financial/operational metrics are more critical drivers.

Model Transparency & Trust

By integrating SHAP visualizations into the final dashboard (discussed later), stakeholders can interactively explore predictions and their drivers. This level of transparency enhances trust, supports model governance, and allows for “what-if” simulations to assess the effect of changing specific input parameters.

7. Streamlit Dashboard Deployment

To make the predictive insights and model outputs actionable and user-friendly, we developed an interactive dashboard using Streamlit, a Python-based web application framework ideal for rapid data app deployment. The dashboard allows supply chain analysts and decision-makers to interact with models in real time, simulate risk scenarios, and extract explainable insights for strategic planning.

7.1. User Input Panel

The left sidebar features a flexible input mechanism where users can manually enter supplier-related features such as:

- On_Time_Delivery_Rate
- Financial_Stability
- Severity_Level

Each input is controlled through intuitive sliders, dropdowns, or numeric fields to ensure ease of use and prevent invalid entries.

7.2. Model Predictions Panel

Upon submission, the dashboard performs real-time predictions for four key KPIs:

- Penalty_Cost_USD (Random Forest)
- Compensation_Paid_USD (Random Forest)
- Disruption_Cost_USD (XGBoost)
- Time_To_Recovery_Days (XGBoost)

Each output is displayed via color-coded metric cards, indicating the predicted value and relative risk level (e.g., low, medium, high). Additionally, a set of bar charts and gauge indicators visually depict the predicted KPI values for quick comprehension.

7.3. Explainability Panel (Optional & Collapsible)

To support transparency, users can opt-in to view SHAP explanations for each prediction:

Local explanations (force plots) are displayed for individual predictions to show how specific inputs influenced the outcome.

This helps build confidence in the model's reasoning and supports deeper diagnosis when unexpected results occur.

7.4. Summary Insights & Recommendations

A dedicated section titled "Risk Summary & Recommendation" summarizes:

- A natural language interpretation of the results (e.g., "High penalty predicted due to low reliability and severe disruption").
- Suggested prescriptive actions, such as exploring alternate suppliers, renegotiating contracts, or monitoring high-risk vendors more closely.
- A risk categorization tag (e.g., Red: High Risk) based on thresholds defined during model evaluation.

The dashboard effectively bridges the gap between machine learning outputs and real-world decision-making. By combining predictions, interpretability, and actionable recommendations in a single interface, it empowers users to manage supply chain risks proactively and confidently.

8. Conclusion

This paper presents a machine learning-based supply chain risk prediction system utilizing **Random Forest** and **XGBoost** models to predict key metrics such as **Penalty_Cost_USD**, **Compensation_Paid_USD**, **Disruption_Cost_USD**, and **Time_To_Recovery_Days**. The models were optimized through **hyperparameter tuning**, and the results are presented through an **interactive Streamlit dashboard**. The system provides valuable insights through **SHAP explainability** and model predictions, helping supply chain managers identify and mitigate potential risks effectively.

In summary, this research highlights the importance of applying **machine learning** to enhance **supply chain risk management** and decision-making, with scope for further improvements in real-time capabilities and model performance.